

# Unsupervised Doppler Radar Based Activity Recognition for e-Healthcare

Karayaneva, Y., Sharifzadeh, S., Li, W., Jing, Y. & Tan, B.

Published PDF deposited in Coventry University's Repository

**Original citation:**

Karayaneva, Y, Sharifzadeh, S, Li, W, Jing, Y & Tan, B 2021, 'Unsupervised Doppler Radar Based Activity Recognition for e-Healthcare', IEEE Access, vol. 9, 9406586, pp. 62984-63001.

<https://dx.doi.org/10.1109/ACCESS.2021.3074088>

DOI 10.1109/ACCESS.2021.3074088

ESSN 2169-3536

Publisher: Institute of Electrical and Electronics Engineers

**This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>**

Received March 18, 2021, accepted April 7, 2021, date of publication April 19, 2021, date of current version April 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074088

# Unsupervised Doppler Radar Based Activity Recognition for e-Healthcare

YORDANKA KARAYANEVA<sup>1</sup>, SARA SHARIFZADEH<sup>1</sup>, WENDA LI<sup>2</sup>,  
YANGUO JING<sup>1</sup>, (Member, IEEE), AND BO TAN<sup>3</sup>, (Member, IEEE)

<sup>1</sup>Faculty of Engineering, Environment and Computing, Coventry University, Coventry CV1 2JH, U.K.

<sup>2</sup>Department of Security and Crime Science, University College London, London WC1E 6BT, U.K.

<sup>3</sup>Faculty of Information Technology and Communication Sciences, Tampere University, 33100 Tampere, Finland

Corresponding author: Bo Tan (bo.tan@tuni.fi)

**ABSTRACT** Passive radio frequency (RF) sensing and monitoring of human daily activities in elderly care homes is an emerging topic. Micro-Doppler radars are an appealing solution considering their non-intrusiveness, deep penetration, and high-distance range. Unsupervised activity recognition using Doppler radar data has not received attention, in spite of its importance in case of unlabelled or poorly labelled activities in real scenarios. This study proposes two unsupervised feature extraction methods for the purpose of human activity monitoring using Doppler-streams. These include a local Discrete Cosine Transform (DCT)-based feature extraction method and a local entropy-based feature extraction method. In addition, a novel application of Convolutional Variational Autoencoder (CVAE) feature extraction is employed for the first time for Doppler radar data. The three feature extraction architectures are compared with the previously used Convolutional Autoencoder (CAE) and linear feature extraction based on Principal Component Analysis (PCA) and 2DPCA. Unsupervised clustering is performed using K-Means and K-Medoids. The results show the superiority of DCT-based method, entropy-based method, and CVAE features compared to CAE, PCA, and 2DPCA, with more than 5%-20% average accuracy. In regards to computation time, the two proposed methods are noticeably much faster than the existing CVAE. Furthermore, for high-dimensional data visualisation, three manifold learning techniques are considered. The methods are compared for the projection of raw data as well as the encoded CVAE features. All three methods show an improved visualisation ability when applied to the encoded CVAE features.

**INDEX TERMS** Activity recognition, data visualization, Doppler radar, health and safety, DCT analysis, unsupervised learning.

## I. INTRODUCTION

Human activity recognition for smart healthcare is an emerging topic. It is becoming even more prominent with the complications of ageing population worldwide. The population aged 65+ in the UK was 11.8 million in 2016, while this number is projected to grow to 20.4 million by 2041 [1]. Chronic and long-term conditions are well-known to increase with age. It is reported that 29% of those aged 60-64 had a chronic condition, while the percentage grows to 50% for elderly populations aged 75 or over. The implications of ageing with chronic conditions prevent elderly people from independent living. Thus, they are dependent on social care services such as living in care homes.

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo<sup>1</sup>.

The demand for human activity detection and monitoring has rapidly increased over the past years. A number of devices are proposed including cameras, wearable technologies, infrared sensors, and radars. These devices are expected to provide daily monitoring of elderly people's activities and vital signs. Hence, this will provide peace of mind for their relatives regarding the physical health and mental health of the care home residents. Cameras are often seen as an obvious and traditional solution for capturing observable data including human activities for subsequent recognition [2]–[4]. Video-depth cameras are capable of obtaining extremely high-resolution data, which can contribute to the detailed analysis of daily human activities. Nevertheless, camera devices suffer from intrusiveness, which is highly undesirable in the contexts of residential environment. The modern healthcare is concerned with the privacy and

dignity of patients. Therefore, vision-based solutions are not recommended in smart care homes.

Wearable sensor technologies are an effective solution for smart healthcare applications as they provide a combination of human activity recognition and vital signs detection [5]–[8]. Wearable sensors have the ability to capture small fractions of the body such as the movement of fingers [9]. Additionally, wearable sensing technologies can detect physiological signals such as heart rate and speech patterns [10]. However, the disadvantages and challenges of this technology are not to be under-rated. Wearable sensors are known to have poor battery life [11]. As they are “wearable”, elderly populations may easily forget to wear the device or feel uncomfortable wearing it [12].

Infrared sensors utilize human’s body temperature distinguished from the lower ambient temperature in order to capture and detect human activities. Most IR sensors obtain ultra low-resolution data, where a subject identification is avoided. Thus, IR devices represent an attractive solution to be deployed in care homes and hospitals. Current studies reveal significant recognition rates ( $>90\%$ ) for activities including standing, sitting, walking, falling, and others [13]–[15]. Contrarily to their advantages, IR devices suffer from a relatively low detection distance. It can be inferred in a recent dataset [16] that the performance drops with distance growth, although not significantly. As IR sensors are low-resolution capturing devices, they lack sensitivity towards small fractions of the human body, which prevents more specific activities detection.

Passive Micro-Doppler radars are an appealing solution for human activity recognition. That is due to their non-intrusiveness, high distance range, deep penetration, and reliable accuracy rates [17]–[19]. In addition, the passive radar uses the existing radio bursts in the environment. It avoids to bring extra RF source to aggravate the increasing electromagnetic interference in the residential environment. While passive Micro-Doppler radars traditionally have applications in human activity recognition [17], they have also been deployed for vitals sign monitoring such as respiration [20]. In addition to their applications, the devices have been used for gait patterns analysis [21].

Micro-Doppler radars have been extensively used for activity recognition with a focus on healthcare purposes [20], [23], [24]. Currently, majority of studies are based on pipelines, which are totally supervised or consist of a combination of unsupervised and supervised approaches. In most cases, the pipelines are based on unsupervised feature extraction methods, such as conventional PCA and Singular Value Decomposition (SVD) techniques. That is usually followed by a supervised classification method, such as Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN) [25], [26]. In the pursuit of a more accurately measured covariance matrix from PCA, variations of PCA have been used for Micro-Doppler data. In [27], the authors applied L1 norm PCA opposed to standard PCA and achieved improved testing accuracies. Furthermore, 2DPCA has been compared with

standard PCA for Doppler radar data [28]. Considering the fact that 2DPCA accepts 2D image matrices as an input, the dependencies of the pixels are retained. The results of that work revealed improved recognition rates for 2DPCA by more than 10%. Furthermore, unsupervised PCA has been combined with supervised Linear Discriminant Analysis (LDA) and shallow neural networks (SNN) [29]. The proposed architecture in that work was the first to use a 3D-signal representation by retaining the matrix dependencies. Results reveal better performance than conventional PCA and 2DPCA for Doppler radar data.

In a pilot study, the *Doppler-Radar-2018* dataset was used. The work employed Hidden Markov Models (HMM) in order to extract activity information from each Doppler sequence [23]. The output of the HMM training was clustered using K-Means and K-Medoids. The Kullback-Leibler (KL) log-likelihood with K-Medoids for clustering obtained the highest accuracy. HMM is a supervised framework that requires the labels and generates log-likelihood values as a measure of similarity of a candidate sample to each of the classes. Therefore, log-likelihood values can be used for decision making directly and the idea of using them as a feature and applying unsupervised methods such as K-Means for clustering them is not the best analysis pipeline. HMM was also used in another previous work [30], for classification of extracted physical features where 72% accuracy was achieved.

Another group of supervised techniques are based on deep learning approaches. These methods require more data for learning their objective functions. Recently, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) architectures have been used. Furthermore, CAE was used for feature extraction for Doppler radar data [31]. That was followed by classification based on a supervised framework by fine-tuning and a Softmax classifier. In these techniques, feature extraction from data streams are performed automatically with minimum user required settings [32].

Unsupervised learning is yet a minimally researched topic for Doppler radar based applications. The advantage of unsupervised methods compared to supervised techniques is that they do not require labeling data. This usually influences the accuracy of unsupervised methods compared to supervised techniques. That is because in the absence of labels, the learning is only guided based on the input variables, their variations and characteristics. That does not necessarily help to learn the decision rules correctly. However, the models can be updated faster compared to the supervised strategies. Hence, the learning capacity of the latter are limited due to labeling requirement for any new coming data. In practical settings, usually recognition of few activities is critical such as a fall or immobility. In future, such activities can be labelled and recognized among the clustered activities. However, this work is only focused on unsupervised activity clustering of Doppler radar data and the latter problem is not addressed in this paper. It will be considered

for future studies. In addition, unsupervised learning usually requires the use of techniques for estimating the number of clusters. This is due to the fact that subjects conduct a broad number of activities in a real world scenario. As such, embedding all activities in a pre-collected dataset for supervised frameworks is problematic.

Unsupervised learning methods can be categorised into two groups of manual or automated feature extraction strategies. In rule-based systems, specially those strategies based on hand-crafted features, prior knowledge about the experimental system and environment are important [33]. Besides that, careful selection of the feature extraction technique and manual selection of some parameters depending on the setup rules are required. For example, the signal strength and angle of measurement can influence filtering window size and scaling or choice of basis function in spatio-temporal feature extraction techniques. On the other hand, in automated feature extraction approaches such as CAEs, such prior settings are not required. The embedded objective function and optimization removes the need for manual settings [34], [35]. Nevertheless, the computational time for automated feature extraction approaches is more expensive compared to manual feature extraction.

In this paper, the *Doppler-Radar-2018* dataset, that was used previously in [23], is considered. An unsupervised framework is developed despite labels availability. As explained earlier, the importance of the designed framework is the applicability to projects with poor labeling scenarios. For this aim, four groups of unsupervised feature extraction strategies are considered: (1) frequency-domain analysis based on 2D DCT (2) entropy analysis (3) convolutional filtering strategies based on CVAE and CAE (4) unsupervised PCA analysis including 1D and 2D analysis. For DCT and entropy feature extraction methods, two methods are proposed. The extracted features are clustered into different activity groups based on unsupervised clustering strategies using K-Means and K-Medoids. In order to evaluate the results, the known labels are utilised only at the result evaluation step. Leave-one-subject-out cross validation (LOOCV) is used so that, the built models are tested on unseen data of one subject. Due to the fact that in unlabelled conditions the number of classes is unknown, four unsupervised metrics, namely Elbow, Silhouette, Davies-Bouldin and Dunn's index are used.

The contributions of the study to the research community are the following:

- **Two proposed unsupervised feature extraction methods for Doppler radar data:** The Doppler radar data in this study has high dimensionality. When reshaped into 2D maps, there are different distinguished patterns for each activity. On the other hand, the high dimensionality leads to an ill-posed problem and over-fitting. Therefore, feature extraction from 2D image maps is employed. For this aim, the local areas with low level of variation and insignificant information can be cancelled out from the analysis. In order to extract the most meaningful

information, two local patching and feature extraction methods are proposed in this paper. To retain the unsupervised scenario and evaluate the features, the Dunn's index is used. It is used as a criterion for evaluation of the activity clustering results, using the locally extracted features. That also helped to choose the first proposed feature extraction method's parameters such as the local patch size and location. The first proposed method uses 2D DCT to extract features from local patches of the 2D images. The second proposed method is based on entropy of the local patches. The average testing results showcase 5%-10% improvement by using the proposed techniques for different scenarios compared to the previous methods. To the best of our knowledge, such unsupervised algorithms have not been used for human activity recognition using Doppler radar data.

- **Comprehensive study of unsupervised learning for Doppler radar data:** This work is a pioneering study concerning unsupervised learning for Doppler radar data. Based on a comprehensive study, four different metrics are used to estimate the number of clusters in the unsupervised framework. That is useful in real scenarios, where the number of activities can be high and recognizing few of them among all clustered activities is required. Additionally in this paper, four groups of unsupervised feature extraction strategies are compared: the proposed methods based on (1) local 2D DCT and (2) local entropy are compared by (3) architectures using deep CVAE and CAE, where the former has not been used previously for Doppler radar data, and (4) previous methods using PCA and 2DPCA. The extracted features are clustered with K-Means and K-Medoids. The proposed methods based on local DCT and local entropy, and CVAE achieved around 5%-20% higher average testing accuracy in comparison with CAE, PCA, and 2DPCA. The proposed methods for feature extraction along with CVAE encoded features can be useful for unsupervised cases or semi-supervised cases with poor labeling.
- **High-dimensional data visualisation enhancement:** Manifold learning methods for high-dimensional data visualisation are considered in this study. Doppler radar dataset can benefit in regards to activities data visualisation. This can reveal similarities between certain activity groups. The manifold learning methods are known to provide good separation between the classes [36]. In this study, CVAE encoded data have been used for data visualisation improvement. For the first time for Doppler radar data, in this research the manifold learning methods' performance over raw data and encoded data using CVAE is employed to illustrate any improvement for the sake of visualisation. The three methods for high-dimensional data visualisation t-Distributed Stochastic Neighbour Embedding (t-SNE), Multidimensional Scaling (MDS) and Locally Linear Embedding (LLE) are compared in the two scenarios.

Initially, the methods are used to transform the raw data  $D_{n \times 6400}$  to a 2-dimensional space. Secondly, the transformation is performed on the encoded features by CVAE data  $D_{n \times 50}$ . Comparison of the results reveal better separations of the clusters using the three methods, when CVAE encoding is used. This showcases the strength of CVAE for data separation. Hence, CVAE encoded features can be used in manifold learning for visualisation purposes.

The rest of the paper is organized as follows: In Section II the methods including the database description, number of clusters estimation and the proposed approaches for feature extraction and clustering are described. Additionally, three methods for data visualisation are defined. The results for unsupervised feature extraction and clustering are shown in Section III. Then, data visualisation techniques are compared by transforming the raw data as well as the CVAE encoded data. Section IV critically evaluates the main findings of the study, including the proposed architectures for unsupervised learning and the manifold learning methods for high-dimensional data visualisation. Finally, Section V concludes the study with the most valuable outcomes.

## II. METHODOLOGY

The unsupervised framework in this study consists of a number of steps as illustrated in Fig. 1. The first step is to divide data into train and test. Five different activities are recorded by Doppler radar. The number of clusters  $K$  is estimated using four metrics. Since the raw data are in high-dimension, unsupervised feature extraction methods are employed to reduce the feature space dimension. The feature extraction is followed by clustering and recognition using K-Means and K-Medoids. Comparison of the proposed two methods for feature extraction - local DCT-based method and local entropy-based method is performed with conventional methods based on CVAE, CAE, PCA, and 2DPCA. The existing CVAE method has not been deployed in previous Doppler radar studies.

### A. DATASET DESCRIPTION

The Doppler-spectrogram dataset is collected in the University of Bristol laboratory. The laboratory experiment layout is shown in Fig. 2 (a 7 m  $\times$  5 m room). The radio source used in this experiment is an Energy Harvesting transmitter (TX91501 POWERCASTER) working on 915 MHz ISM band with 30 dBm DSSS signal. The passive radar is a two-channel software defined radio (SDR), which is built on two synchronized NI USRP 2920s. Both channels are connected with directional antennas. The reference channel is 1 m apart from the transmitter, while the surveillance channel is pointed to the subject. The Cross-Ambiguity Function (CAF) which is the Fourier of cross-correlated reference and surveillance signals is used to 2D range-Doppler plot. From each range-Doppler plot, the range column which contains the detected subject is extracted to form the Doppler

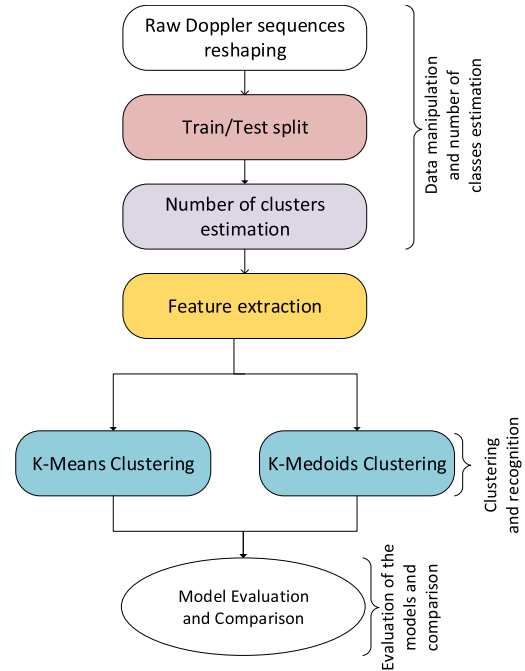


FIGURE 1. Flowchart, describing the overall analysis framework of the paper.

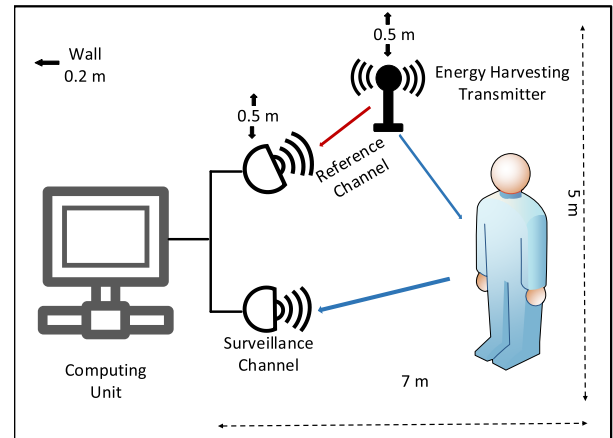


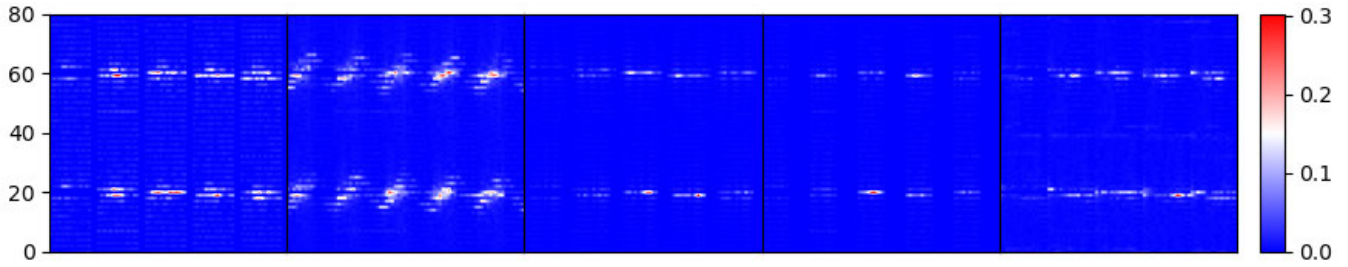
FIGURE 2. Experimental layout.

spectrogram. More details can be found in Section 3 in [20] or Section III in [23].

Four participants (one female and three male) volunteered for capturing activities. This dataset consists of five activities: (1) walking, (2) running, (3) jumping, (4) turning, and (5) standing. Each activity is repeated 10 times by each subject. There exist 40 samples for each activity or 200 samples totally.

A pre-processed Doppler radar dataset is used in this study. The total number of features per sample is  $6400 = (2 \text{ directions} \times 100 \text{ Doppler bins} \times 32 \text{ time index})$ . Furthermore, the Doppler radar data is normalized, which corresponds to the fact that all features are represented by real values in the range of (0, 1). Considering the 3-dimensionality of the Doppler radar data, it is then vectorized  $2 \times 100 \times 32$ , which results





**FIGURE 3.** Example of an  $80 \times 80 = 6400$  image for each activity: (1) walking, (2) running, (3) jumping, (4) turning, and (5) standing.

in 6400. In order to transform it to 2D maps,  $80 \times 80$  reshaping is applied. The reason for converting the vectorized Doppler radar data into 2D maps, is to apply image analysis strategies for quantification of local variation and patterns in the image. Fig. 3 is the micro Doppler signature for human activities.

The Python libraries used for data pre-processing are pandas (version 1.0.5) and numpy (version 1.19.1). In terms of machine learning for feature extraction and clustering, scikit-learn (version 0.22.2) and scikit-learn-extra (version 0.1.0b2) modules are applied. The CAE and CVAE are implemented and run with Keras (version 2.2.4) and Tensorflow (version 2.2.0). The visualisation results are implemented with matplotlib (version 3.2.2).

## B. NUMBER OF CLASSES ESTIMATION

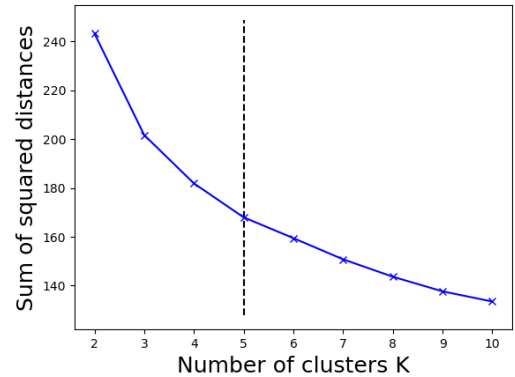
Considering the unsupervised scenario in this study, the number of classes/clusters is unknown. In order to estimate the correct number, a number of techniques are applied including Elbow method, Silhouette analysis, Davies-Bouldin score and Dunn's index using K-Means clustering.

### 1) ELBOW METHOD

The Elbow method is a heuristic technique for clusters number estimation [37], [38]. The overall goal for the method is to maximize the inter-class variability and minimize the intra-class variability. In this study, the data samples are denoted as  $D = D_1, D_2, \dots, D_n$ . The number of clusters is  $K$  and their centroids are given by  $\omega_1, \omega_2, \dots, \omega_K$ . The distortion  $J$  is used to measure the effectiveness of the method:

$$J(K, \omega) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^K (\min(D_i - \omega_j)^2) \quad (1)$$

In this study, the candidate numbers of clusters  $K = 2, 3, \dots, 10$  are selected for K-Means clustering, which is described later in this section. The Elbow method computes the sum of squared errors for the data samples in each cluster. As the number of clusters increases,  $J$  becomes smaller. However, the best value of  $J$  is the point, where a further increase to the number of clusters does not change the within-cluster sum of squares significantly. However, a further increase would result in over-clustering. The decrease trend of  $J$  is noticeable before reaching the actual number of clusters  $K$  and becomes smoother afterwards. The Elbow method is a visualisation tool, and the graph shows a noticeable



**FIGURE 4.** An Elbow test used to determine the number of clusters  $K$ .

decline when the curve approaches the actual  $K$ . Therefore, the decline becomes smoother after exceeding  $K$ . Fig. 4 illustrates the Elbow test for this data.

As it can be observed, the selected number of clusters  $K = 5$ , which is the actual number of clusters for this study. However, detection of this bend point is ambiguous in some cases. Therefore, additional techniques are considered in this study.

### 2) SILHOUETTE ANALYSIS

Silhouette analysis is one of the most commonly used techniques for number of clusters estimation [39]. The method is given as:

$$Silhouette = \frac{1}{n} \sum_{i=1}^n \frac{b(D_i) - a(D_i)}{\max\{a(D_i), b(D_i)\}} \quad (2)$$

where  $a(D_i)$  is the average distance between data point  $D_i$  and the remaining data points in its own cluster. The minimum average distance between data point  $D_i$  and all other clusters is denoted with  $b(D_i)$ . The Silhouette coefficient aims to show the suitability for data point  $D_i$  to belong to a particular cluster. The score is within the range of  $(-1, 1)$ , where a lower value refers to overlapping clusters. On the other hand, a higher value suggests well-separated clusters.

### 3) DAVIES-BOULDIN INDEX

The Davies-Bouldin index is a clusters estimation method concerned with identifying clusters, which are distinct from

each other [40]. The measure is given by:

$$Davies = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} \frac{\Delta(C_i) + \Delta(C_j)}{\lambda(\omega_i, \omega_j)} \quad (3)$$

where  $\Delta(C_i) = \sum_{D_i \in C_i}^n \|D_i - \omega_i\|^2$ , which is the distance between each Doppler sequence and the centroid of the corresponding cluster. The distance between cluster centroids is given by  $\lambda(\omega_i, \omega_j) = \|\omega_i - \omega_j\|^2$ . Then, the term for maximization is computing the ratio of within-cluster to between-cluster distances for the  $i^{th}$  and the  $j^{th}$  clusters. This is computed for all combinations of cluster  $i$  and other clusters. Then, the maximum value is found for each  $i$ . Among all combinations of clusters, the closest clusters with largest spreads have the maximum ratio. That is the worst scenario. The desire is to minimize the overall average of the worst scenario ratios. Therefore, unlike previous methods, a smaller value for Davies index is desirable.

#### 4) DUNN'S INDEX

Dunn's index is one of the most popular and oldest techniques in the literature [41] for number of clusters estimation. The overall aim is to minimize the intra-cluster distance and maximize the inter-cluster distance. It is given by:

$$Dunn = \frac{\min_{1 \leq i < j \leq K} \lambda(\omega_i, \omega_j)}{\max_{1 \leq k \leq K} \Delta(C_k)} \quad (4)$$

where  $\lambda(\omega_i, \omega_j)$  is the distance between clusters  $C_i$  and  $C_j$ . The intra-cluster distance of a single cluster is given by  $\Delta(C_k)$ . The minimization in the numerator finds the Euclidean distance of the two closest clusters. On the other hand, the maximisation in the denominator, finds the Euclidean distance of the samples to the centroid of the cluster with the highest dispersion. Therefore, for optimum cluster number, the numerator will be the maximum value among all other candidate number of clusters, while the denominator will be the lowest, resulting a peak over the heuristic search.

Similarly to the Elbow method implementation, a candidate set of clusters is given  $K = 2, 3, \dots, 10$  for K-Means. The heatmap in Fig. 5 shows the results for each heuristic technique. It is important to note that the value for the inverse of Davies index is used. Hence, the highest values corresponding to the darkest colors show the best fit for the clusters. The selected number of clusters is  $K = 5$  for all methods.

### C. FEATURE EXTRACTION METHODS

Traditionally, researchers rely on the empirical knowledge for the human activity recognition using micro-Doppler signatures. The typical examples include average torso velocity, period or duration of the activity cycle, upper and lower envelope variances in [25], [42]. The empirical knowledge presents intuitive relation between the Doppler signature and feature, however, it is not suitable for

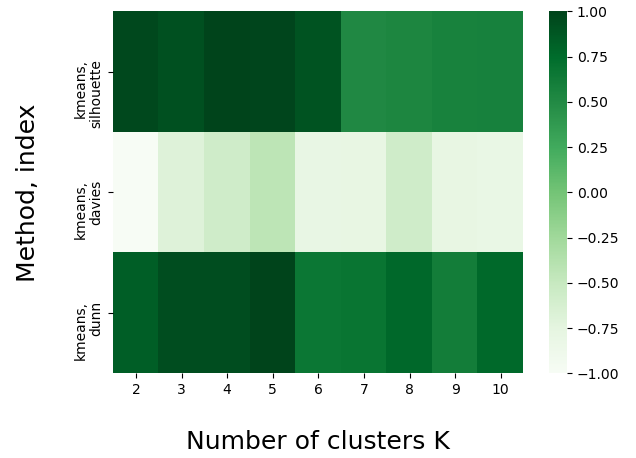


FIGURE 5. Silhouette score, Davies score and Dunn's index used to determine the number of clusters  $K$ .

the datasets with uncontrolled or unknown conditions. Later on, the data-driven micro-Doppler recognition approaches are proposed and proven excellent performance in [43]–[45]. These approaches treat the micro-Doppler plots as time-spectrogram and range-Doppler time points cluster respectively. In this work, we will explore two new local DCT-based and local entropy-based methods as well as convolutional filter-based and variation-based projection methods for feature extraction. The two new methods and convolutional filter-based strategies are superior in terms of accuracy. While the convolutional strategy is superior in accuracy, the training time of the new methods is considerably less than the convolutional strategy.

#### 1) THE PROPOSED LOCAL DCT-BASED METHOD

The first proposed method for feature extraction is based on applying 2D DCT on local areas of the 2D map of Doppler radar data. In general, 2D DCT is used to transform 2D images from the spatial domain to the frequency domain [46]. The sharp changes or smooth variations in the images correspond to high frequencies or low frequencies in DCT domain respectively. In fact, the frequency information of the images is sorted by DCT transform. The DCT coefficient in the top-left corner of the output 2D DCT matrix corresponds to the lowest frequency of zero and the frequencies increase toward the bottom-right corner. Depending on the type of images, the higher energies appear in different coefficients. The most valuable information is usually in small fractions of the DCT images. Therefore, DCT allows selection of a limited number of features that reduces the dimension of the feature space. The method has been used previously for extracting features from micro-Doppler radar for human activity recognition [47]. The mathematical notation of DCT is given as:

$$F[p, q] = \frac{2}{\sqrt{uv}} a_p a_q \sum_{i=0}^{u-1} \sum_{j=0}^{v-1} f[i, j] \cos \frac{\pi(2i+1)p}{2u} \cos \frac{\pi(2j+1)q}{2v} \quad (5)$$

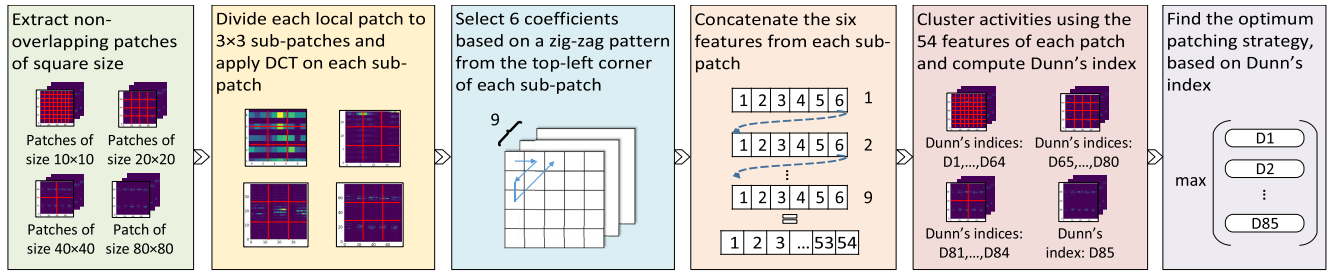


FIGURE 6. Flowchart of the proposed local DCT-based method.

where,  $F[p, q]$  values are the DCT coefficients at row  $p$  and column  $q$ . In addition,  $f[i, j]$  is the element in row  $i$  and column  $j$  of the image matrix, where  $u = 80$  and  $v = 80$ , and  $0 \leq p \leq u - 1$  and  $0 \leq q \leq v - 1$ . Moreover,  $a_p = \frac{1}{\sqrt{2}}$  if  $p = 0$  and it is 1 otherwise. Similarly,  $a_q = \frac{1}{\sqrt{2}}$  if  $q = 0$  and its value is 1 otherwise.

Considering the local variations of the original  $80 \times 80$  images, in this paper, a systematic search algorithm is proposed in order to find the best strategy for applying 2D DCT. This includes applying DCT on different local areas of the  $80 \times 80$  2D maps, using various patch sizes. The aim is to identify the optimum patch size giving the best clustering results. As illustrated in Fig. 6, first the 2D images are divided into various square shape local patches of different sizes. The square-sized local patches are non-overlapping. Four sizes of local patches are considered -  $10 \times 10$ ,  $20 \times 20$ ,  $40 \times 40$  and the original  $80 \times 80$  2D map. Second, each patch is divided into  $3 \times 3$  sub-patches. Third, 2D DCT is applied to each local sub-patch and the resulting 2D map of the DC coefficient's amplitude is used for feature selection. Third, six coefficients are extracted according to a zig-zag pattern from the top-left corner of each sub-patch, allowing to extract features from all local areas of the Doppler profiles. Finally, the six features from each of the nine sub-patches are concatenated to form a feature vector of size  $9 \times 6 = 54$ . This generates 54 features for each local patch. Then, the activities are clustered using the 54 features of each patch and the Dunn's index is computed. The highest Dunn's index discovers the most optimum local patch for the 2D DCT analysis.

The reason for using the Dunn's index in this algorithm is that it describes the quality of the resulting clusters. It quantifies an easily interpretable metric based on the worst clusters of a clustering scenario. As shown in (4), in its numerator there is the minimum between-cluster distance and the denominator is the maximum within-cluster distance. Then, a high Dunn's index shows a good clustering quality. The use of Dunn's criterion rather than clustering accuracy allows parameter selection for the unsupervised framework.

## 2) THE PROPOSED LOCAL ENTROPY-BASED METHOD

Considering the  $80 \times 80$  images, a texture analysis method based on entropy is proposed to quantify the patterns of different activities profiles. Entropy is a statistical measure of randomness and is formulated based on Shannon's equation

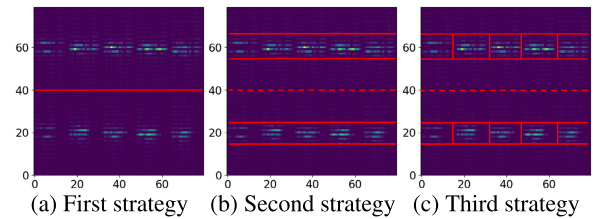


FIGURE 7. The three patching strategies for entropy analysis.

[48] as follows:

$$H(\rho) = - \sum_{i=1}^b \rho_i \log(\rho_i) \quad (6)$$

where  $\rho_i$  is the normalized histogram counts. It is calculated based on the histogram of the image.  $b$  is the total number of histogram bins.

Depending on the variations of colors in local image area, the entropy can change. If most pixels in an image are similar with a low level of variations, the entropy will be small. On the other hand, if the level of color variation is high in an image, the entropy increases. Therefore, depending on the location of the analysis window, the entropy value can change. Since the patterns and color intensities vary for different activities, a careful selection of local patches can generate different entropy values suitable for discrimination of the activities. Based on the observed changes in Fig. 3, three patching strategies are considered so that, the selected image areas for entropy analysis are narrowed down systematically. The three local patching strategies are illustrated in Fig. 7. Then, similar to the local DCT-based analysis, the Dunn's index is used to evaluate the quality of the clustering results based on the entropy features. That allows identifying the best patching strategy. The steps of the proposed method are outlined in Fig. 8.

## 3) CONVOLUTION FILTER-BASED METHODS

In this section, a description of CAE is provided by considering its drawbacks. Then, CVAE is introduced, which overcomes the drawbacks of the previous architecture.

### a: CONVOLUTIONAL AUTOENCODER (CAE)

Autoencoders (AEs) are unsupervised neural networks, which can be used for feature extraction. Their architecture consists of two components: an encoder and a decoder [49]. AEs are commonly used for data denoising [50], anomaly



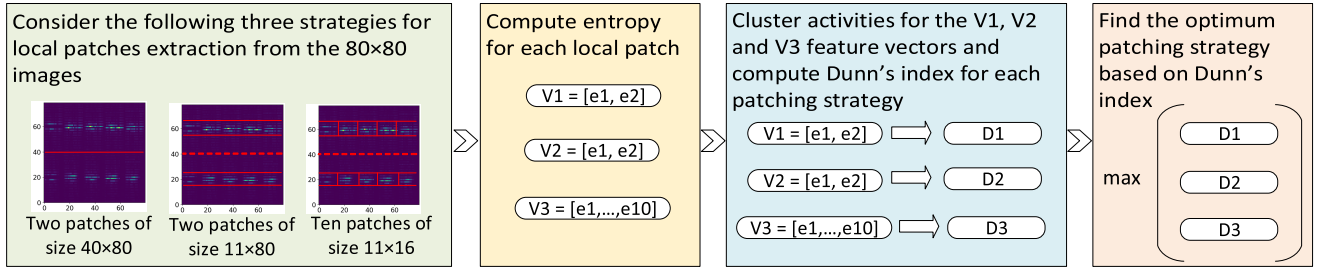


FIGURE 8. Flowchart of the proposed local entropy-based method.

detection [51] and image generation [52]. The encoder learns the latent attributes of the input data  $x$  and transforms it to a lower dimensionality representation  $z$ . On the other hand, the decoder aims to reconstruct  $x$  given  $z$ . The implementation of CAE only contains a reconstruction loss, which needs to be minimized.

In regards to disadvantages of the discussed architecture, the CAE learns local parameters for each data point. This avoids any statistical strength to be shared across all data points. Hence, this may result in overfitting due to the inability of the model to generalise. In addition, the CAE architecture includes only a reconstruction loss and it lacks any regularisation term as seen in CVAEs. This leads to data points of the same group/class to be given different representations, which are often meaningless.

In this work, a deep CAE is used with three hidden layers for the encoder and decoder as illustrated in Fig. 9. As it can be seen, the shape of the input data  $x$  is  $2 \times 100 \times 32$  and the retained number of latent variables  $z$  is 50. The structure of the encoder is symmetric to the decoder's structure. In regards to the hidden layers, the first convolutional layer in the encoder and the third convolutional layer in the decoder have 256 filters with size  $2 \times 3$ . The stride for these two layers is (1, 2) referring to height and width. The second convolutional layer in the encoder and the decoder have 128 filters with size  $1 \times 3$  and the stride is of shape (1, 2). Considering the third layer in the encoder and the first layer in the decoder, they have 64 filters with size  $1 \times 3$ . Their stride is of shape (2, 1). The robustness of CAE is validated in Section III. The convolutional layers used for this architecture incorporate a ReLU activation function. The decoder's task is to reconstruct  $x$  given  $z$ , which is evaluated with the reconstruction loss. Based on this architecture, the encoded features  $z$  are used for clustering the activities.

#### b: CONVOLUTIONAL VARIATIONAL AUTOENCODER (CVAE)

CVAEs are generative models defined in [53], which are commonly used for dimensionality reduction [54], data augmentation [55], and reinforcement learning [56]. Considering Doppler radar data, CVAEs have been used for synthetic data generation [57]. CVAEs contain two main modules: an encoder, referred to as recognition model or inference model, and a decoder, also defined as generative model [58]. The purpose of the encoder is to learn the stochastic mappings of the observed input space  $x$  with a rather

complicated distribution and transform it from its original high-dimensional space (6400 dimensions in this case) into a much lower latent representation  $z$  with a relatively simple distribution. Then, the output of the recognition model  $z$  is the input of the decoder. The decoder aims to reconstruct the original input  $x$  from the reduced latent information. In a traditional autoencoder, the latent representation consists of single-valued outputs for each feature. CVAEs assume that the dimensions of  $z$  cannot be interpreted with simple variables. Instead, CVAEs introduce a probability distribution for the samples of  $z$ , which is commonly a Gaussian distribution [59]. The use of a single reconstruction error in encoder-decoders might result in encoding some meaningless content. That results in overfitting and therefore the latent space should be regularised. Contrarily to CAE, in the CVAE architecture, the loss function includes a reconstruction term and a "regularisation" term. The latter term is developed by enforcing the probabilistic distribution of the encoded space to be close to a Standard Normal distribution. This is expressed as the Kullback-Leibler (KL) divergence. The KL divergence quantifies the divergence of the latent space distribution, denoted as  $q_\theta(z|x)$  in (7) and the standard normal distribution  $p(z)$ :

$$l_i = -E_{z \sim q_\theta(z|x_i)}[\log p_\phi(x_i|z)] + KL(q_\theta(z|x_i)||p(z)) \quad (7)$$

where  $p_\phi(x_i|z)$  describes the generative probability of the reconstructed output given the encoded variable  $z$ . The distribution of the encoded variable  $z$  given the input  $x_i$  is denoted as  $q_\theta(z|x_i)$ . In addition,  $\theta$  and  $\phi$  are parameters of the distribution.

The architecture of the developed CVAE model is illustrated in Fig. 10. Similar to the CAE architecture, the input for CVAE is of shape  $(2 \times 100 \times 32)$  corresponding to height, width, and depth. The ReLU activation function is also used similarly. The structure of the encoder is again symmetric to the decoder's structure. The first convolutional layer in the encoder and the second convolutional layer in the decoder have 128 filters with size  $1 \times 3$ . Their stride is of shape (1, 2) corresponding to height and width. The second layer in the encoder and the first layer in the decoder have 64 filters with size  $1 \times 3$ . Their stride is of shape (2, 2). The encoding part does not forward the direct latent values to the decoder. Instead, mean  $\mu_\phi$  and variance  $\log \sigma_\phi^2$  vectors of the latent features are its output. These parameters

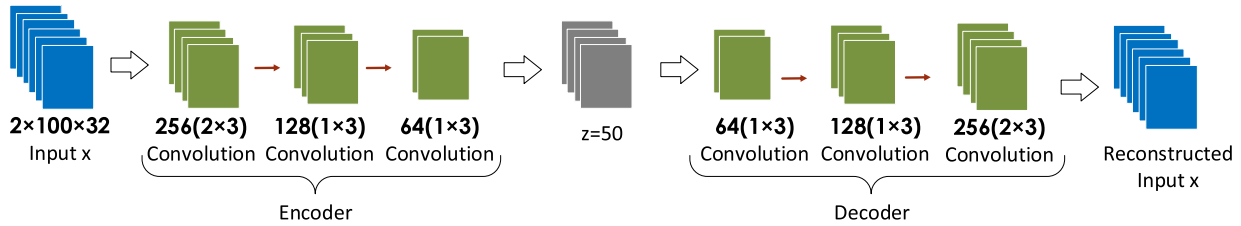


FIGURE 9. The CAE architecture.

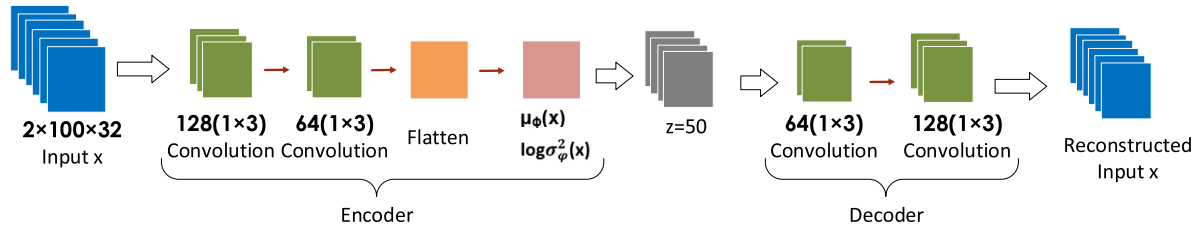


FIGURE 10. The CVAE architecture.

are enforced to be close to a standard normal distribution, which is measured by the regularisation term. Finally, both  $\mu_\phi$  and  $\log \sigma_\phi^2 x$  are sampled to produce the compressed latent space representation  $z$  with the specified number of features. Considering the latent representation  $z$ , the decoder aims to reconstruct  $x$ . The effectiveness of this operation is evaluated with the reconstruction loss.

#### 4) VARIATION-BASED PROJECTION TECHNIQUES

##### a: PRINCIPAL COMPONENTS ANALYSIS (PCA)

1D-PCA is known to be one of the most common linear techniques for unsupervised feature extraction [60]. PCA finds the directions of main variations of data in the original high dimensional space, and projects data along those directions into a smaller sub-space. Based on this linear projection, the dimensionality of data is reduced. In this paper, the Doppler radar data with 6400 variables are used for PCA analysis. Based on a weighted linear combination of these features, the main directions of variations of data are calculated. In this paper, the  $s$  number of the first few eigen vectors, explaining 95% of data variations, is used to transform the original high dimensional data  $D_{n \times m} W_{m \times s} = Z_{n \times s}$ . The first principal component (PC1) usually retains the highest variance, which allows a smaller number of PCs to be selected.

##### b: 2D PRINCIPAL COMPONENTS ANALYSIS (2DPCA)

PCA requires the 2D image matrix to be transformed to a 1D image vector. This often leads to a high-dimensional image vector. Therefore, the size of the covariance matrix  $V$  is extremely large. Logically, it becomes difficult to evaluate the covariance matrix considering the small number of training examples. 2-dimensional PCA proposed in [61] allows the covariance matrix to be calculated on the 2D images of size  $80 \times 80$ . Hence, this corresponds to its smaller size, which has two main advantages. Less computation time is required and the covariance matrix is more accurately evaluated.

The covariance matrix is given by:

$$V = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^T (X_i - \bar{X}) \quad (8)$$

where  $n$  is the number of training samples and  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  is the average training image. It also has the size  $80 \times 80$ .

More specifically, 2DPCA computes the covariance matrix only for the row or column dimension only. That is because the initial data is not vectorized to include all features. Considering the  $80 \times 80$  2D images in this study, only the columns are used for computing the covariance matrix. This process is followed by eigen decomposition of the covariance matrix. PCs retaining most of the variance are then selected. Similar to PCA, the first PC retains most of the variance.

#### D. CLUSTERING METHODS

Two clustering methods are used for grouping the data samples. Both of the methods are distance-based: K-Means and K-Medoids.

##### 1) K-MEANS

K-Means clustering is one of the most commonly used techniques for unsupervised learning [62]. The K-Means method considers the number of groups or clusters  $K$  is known and it aims to group the data points based on their distances. In this work, the Euclidean distance is used. The overall goal for the clustering method is to group the data points  $D_1, D_2, \dots, D_n$  in clusters by minimizing the intra-cluster distances. Intuitively, the distances between data points from different clusters should be maximized. During training, K-Means outputs the cluster centres  $\omega_k$ , where  $k = 1, 2, \dots, K$ , or also known as centroids. The assignment of a new data point to one of the clusters is such that, the sum of the squared distances between the data point and all cluster centroids  $\omega_1, \omega_2, \dots, \omega_k$  are computed. Then, the sample is assigned to the cluster, where the corresponding distance is minimum.

The objective function for K-Means, which specifies the sum of squared distances of each data point  $D_i$  to cluster  $k$ , is defined as follows:

$$J = \sum_{i=1}^n \sum_{k=1}^K r_{ik} \|D_i - \omega_k\|^2 \quad (9)$$

where  $r_{ik}$  is a binary function indicating the assignment of data point  $D_i$  to cluster  $k$ . If  $D_i$  is assigned to cluster  $k$ , the binary indicator  $r_{ik} = 1$ , and 0 otherwise:

$$r_{ik} = \begin{cases} 1, & \text{if } k = \arg \min_j \|D_i - \omega_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The overall aim is to minimize the  $J$  for values  $r_{ik}$  and  $\omega_k$ . The procedure can be achieved by iterative optimization with respect to  $r_{ik}$  and  $\omega_k$ . In the first phase, the  $\omega_k$  is fixed, while the goal is to optimize  $r_{ik}$ . The same notion is applied to the second step as  $r_{ik}$  is fixed and the focus is on the optimization of  $\omega_k$ . The entire process corresponding to Expectation-Maximization algorithm is repeated until convergence.

## 2) K-MEDOIDS

K-Medoids is a clustering method based on distances analysis, which has shown better performance for noisy and problematic data than K-Means [63]. Similar to K-Means, this method considers the number of groupings or clusters  $K$  is known initially and  $K < n$ , where  $n$  is the number of data points. In contrast, K-Medoids considers a data sample for the centroid or the medoid, which is not the case for K-Means. In the first step of K-Medoids, the algorithm aims to find a data point  $D_i$  in a cluster  $C(i) = k$  that is in minimum distance to the remaining observations in the cluster  $D'_i$ . This distance is denoted as  $\|D_i - D'_i\|^2$  and the minimisation is shown in (11):

$$i_k^* = \arg \min_{i: C(i)=k} \sum_{C(i)=k} \|D_i - D'_i\|^2 \quad (11)$$

Then, the output index  $i_k^*$  is used to find a new centroid or medoid, defined as  $\omega_k = D_{i_k^*}$ ,  $k = 1, 2, \dots, K$  for all clusters.

The second step of the method is to minimise the total error by re-assigning each data sample to the closest centroid. The clusters centroids are given  $\omega_1, \omega_2, \dots, \omega_K$ .

$$C(i) = \arg \min_{1 \leq k \leq K} \|D_i - \omega_k\|^2 \quad (12)$$

Finally, step 1 and step 2 are iterated until the algorithm converges to the optimum centroids.

## E. VISUALISATION TECHNIQUES FOR HIGH-DIMENSIONAL DATA

In this paragraph, three widely-known techniques for manifold learning are defined. The focus is on transforming the very high-dimensional space in this dataset ( $D_{n \times m} = D_{n \times 6400}$ ) to a 2-dimensional space. Manifold learning methods are known to map closely correlated data samples in

similar positions, while the gap in the low dimensional space increases if the samples are non-similar. Hence, comparison measures will be extracted from this analysis, which can be useful for projects concerned about high-dimensional data visualisation.

### 1) T-DISTRIBUTED STOCHASTIC NEIGHBOUR EMBEDDING (T-SNE)

T-SNE is a non-linear dimensionality reduction method, which has gained attention for its superior ability to visualise high-dimensional data by transforming it to a two or three-dimensional space [64]. This method assigns each data point in a low-dimensional location by aiming to preserve the significance of the original information. Unlike linear techniques such as PCA and SVD, t-SNE aims to keep similar data points in close locations in the low-dimensional space. The superiority of t-SNE in comparison with other dimensionality reduction methods is the ability to preserve the local structure of the data as well as global information such as clusters. T-SNE has been recently compared with PCA for visualisation where the former achieved better visualisation [65]. The steps for the t-SNE transformation are described below:

- 1) The Doppler sequences  $D_1, D_2, \dots, D_n$  are initially in their original 6400-dimensional space. T-SNE begins with determining the similarity between the data samples. This is performed by computing their distances. Euclidean distances are used in this study.
- 2) The Euclidean distances are converted to probabilities describing normal distributions so that, similar data samples have close values. On the other hand, dissimilar points have distinct similarity values. The similarity scores are calculated for each data points pair  $D_{ij}$ , where a similarity matrix is obtained based on probabilities  $p_{ij}$ .
- 3) The data samples are projected in a random order to the low dimensional space first. This results in a mismatch with cluster patterns of data in the original domain initially. The aim for t-SNE is to re-position the data samples in the new low dimensional space, such that the same clustering patterns of the high-dimensional space to be preserved.
- 4) Then, the Euclidean distances between the data samples are calculated in the lower dimensional space. Similarly, the distances are converted to a t-Distribution (e.g.  $q_{ij}$  for the two data points  $D_i$  and  $D_j$ ). t-Distribution is similar to normal distribution, but with taller tails. The taller tails of t-Distribution prevent dissimilar data points to be positioned in close locations of the lower dimensional space. The samples in lower dimension are re-positioned using these probabilities resembling distances. The re-positioning is performed based on the two probabilities  $q$  of low dimension and  $p$  of high dimension.
- 5) t-SNE uses KL divergence to optimise the similarity of the distributions described by  $q$  to those described

by  $p$ . This can be interpreted as a constant comparison of the samples distances in the lower-dimension to their distances in the original high-dimension. Then, re-positioning will be improved iteratively, as the similarity matrix of probabilities  $q$  is optimized using the original similarity matrix  $p$ .

## 2) MULTIDIMENSIONAL SCALING (MDS)

MDS is a non-linear dimensionality reduction technique. It can be used for visualization of high dimensional data in low dimensional space. It preserves the actual distances of original samples in the low dimensional space. MDS considers dissimilarities of sample pairs contrary to other methods, which are concerned with similarities. Given the set of observation  $D_1, D_2, \dots, D_n \in \mathbb{R}^m$ ,  $d_{ij}$  is the dissimilarities e.g. the Euclidean distance of two samples  $D_i$  and  $D_j$  so that,  $d_{ij} = ||D_i - D_j||^2$ . MDS seeks  $z_1, z_2, \dots, z_n \in \mathbb{R}^k$ , so that  $k < m$ . A so called stress function is minimized for this aim [66]–[68]:

$$\text{Stress}(z_1, \dots, z_n) = \sum_{i \neq j}^n (d_{ij} - ||z_i - z_j||)^2 \quad (13)$$

where  $||z_i - z_j||$  is the Euclidean distance between  $z_i$  and  $z_j$ . Then the pairwise distances are preserved in the lower dimensional representation. A gradient descent algorithm is used to minimize the stress function and find the components in the low dimension [68]. MDS transformation is monotone increasing with the increasing dissimilarities. The same notion is applied to growing similarities data, which decreases the transformation. Hence, similar object pairs are positioned closely in the transformed space, while objects with dissimilarity are distinguished with larger distances.

## 3) LOCALLY LINEAR EMBEDDING (LLE)

LLE is a non-linear dimensionality reduction method proposed in [69]. The method is concerned with preserving the global structure of the data based on an underlying manifold. The data are represented by  $n$  real-valued vectors  $\vec{X}_i$  in a high dimension  $m$ .  $i$  is the index of a sample. Each data point  $\vec{X}_i$ , is a member of a neighbourhood. Each neighbourhood consists of similar data points. Similar data points are expected to lie on a close locally linear patch of the smooth manifold. The  $p$  nearest neighbours for each data point are defined by measuring the Euclidean distances. The local geometry of the patches can be characterised by linear coefficients. These linear coefficients are used to reconstruct each data point from its neighbours. The reconstruction loss is defined by:

$$\epsilon(W) = \sum_{i=1}^n ||\vec{X}_i - \sum_{j=1}^p W_{ij} \vec{X}_j||^2 \quad (14)$$

where  $W_{ij}$  are the weights defined for data points reconstruction using the corresponding neighbours. The number of samples is given as  $n$ , while the number of neighbours is  $p$ . The computed weights  $W_{ij}$  correspond to the contribution of a data point  $\vec{X}_j$  for reconstructing  $\vec{X}_i$ . In order to ensure that  $\vec{X}_i$

is reconstructed only by its neighbours, the weight function  $W_{ij} = 0$ , if a sample  $\vec{X}_j$  does not belong to the same class. Another constraint to the reconstruction loss is that the sum of the weight matrix's rows should be one,  $\sum_{j=1}^p W_{ij} = 1$ .

This sum-to-one constraint makes the weights invariant to translation of the data points and their neighbors. The weights are also invariant to rotation and scaling. The minimisation of the loss function, allows computation of the weights  $W$ . They characterize the intrinsic geometric properties of each neighborhood.

Using the weights,  $W_{ij}$ , it is possible to project each high-dimensional data point  $\vec{X}_i$  to vector  $\vec{Y}_i$  of the lower representation based on another reconstruction cost function. Having the  $W_{ij}$  fix, the aim is to minimise the embedded cost function to optimise the low  $d$ -dimensional coordinates ( $d < m$ ):

$$\Phi(Y) = \sum_{i=1}^n ||\vec{Y}_i - \sum_{j=1}^n W_{ij} \vec{Y}_j||^2 \quad (15)$$

## III. EVALUATION AND RESULTS

In this section, the results obtained using the four groups of unsupervised feature extraction techniques are presented. These include the two proposed methods, namely, local DCT-based method and local entropy-based method. In addition, the existing convolutional filter-based, and variational-based projection methods are used for comparison.

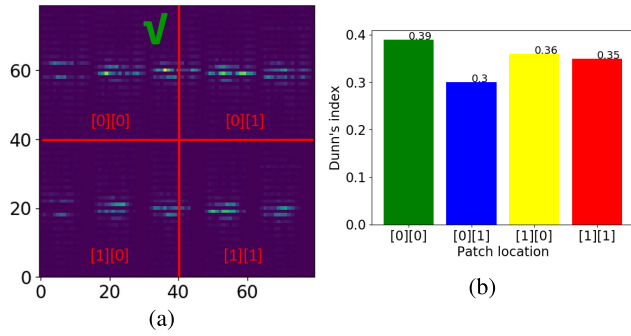
For local DCT-based method, local entropy-based method, and 2DPCA, the inputs are  $n$  reshaped images of size  $80 \times 80$ . In the case of CAE and CVAE, the inputs are  $n$  number of 3D cubes of size  $2 \times 100 \times 32$ . While for PCA, the input data is  $D_n \times 6400$ . In the case of PCA and 2DPCA, the selected number of eigen vectors preserves 95% of the data variance.

Leave-one-subject-out cross validation (LOOCV) is used to avoid over-fitting. As listed in Section II, four participants are included in the data. The four participants correspond to the four folds. The models are trained on three subjects data. Then, they are validated on unseen data from the remaining subject, which is not used for building the models. This is repeated for all four subjects. Hence,  $Z_{tr} = 150 \times m$  is the training matrix and  $Z_{ts} = 50 \times m$  is the matrix for testing, where the number of features  $m$  varies for different models.

In addition to the activities clustering results, the three manifold learning methods t-SNE, MDS and LLE are compared in two scenarios. In the first scenario, they are used to transform the raw data features to a 2-dimensional space. Since CVAE encoded features obtained the most accurate clustering results, in the second visualization scenario, they are used for projection into a 2-dimensional space.

True labels are only used for model evaluation and illustration purposes. The order of the predicted labels by clustering is not necessarily consistent with the actual labels order. Therefore, the clustering accuracy is estimated by finding the best-matching pairs of clusters labels and true labels. Based





**FIGURE 11.** (a) Illustration of the  $40 \times 40$  local patches in the Doppler spectrogram. The optimum patch for feature extraction is ticked in green color and (b) The corresponding Dunn's index for the four  $40 \times 40$  local patches.

**TABLE 1.** Average and standard deviations of testing accuracies based on K-Means and K-Medoids using DCT features from the raw data  $80 \times 80$  single patch and those from the selected  $40 \times 40$  local patch features over 4 rounds of LOOCV.

	DCT Raw Data	DCT-Based Method
K-Means	$63.5\% \pm 8.64$	$75\% \pm 5.74$
K-Medoids	$62\% \pm 9.89$	$77\% \pm 4.58$

on this, the predicted labels by clustering are matched to their corresponding actual true labels. As such, a regular accuracy score function is used for calculating the accuracy.

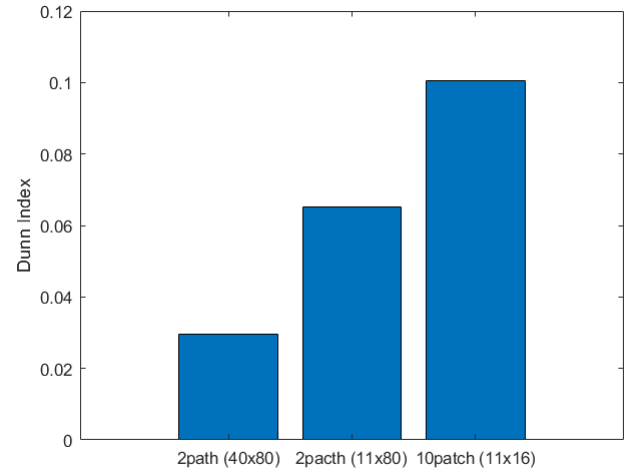
### A. LOCAL DCT-BASED ANALYSIS RESULTS

The proposed local DCT-based method extracts non-overlapping square-sized patches from the original  $80 \times 80$  2D map for analysis. Dunn's index is used for validating the method, which showed the highest values for the  $40 \times 40$  local patches. In Fig. 11a, the four possible non-overlapping patch locations for this size are illustrated. Then, each  $40 \times 40$  patch is divided into  $3 \times 3$  sub-patches as was shown in Fig. 6 previously. Next, six DCT coefficients of the top-left zig-zag pattern are selected from each sub-patch yielding a total of  $9 \times 6 = 54$  DCT coefficients for each  $40 \times 40$  patch. The features are then used for clustering. The resulting Dunn's indices are visualized and compared in Fig. 11b. As observed, the  $40 \times 40$  patch in the top-left corner of the 2D image is found as the best location in terms of Dunn's index for DCT analysis. That shows the lower order frequencies coefficients are related to detection of activities.

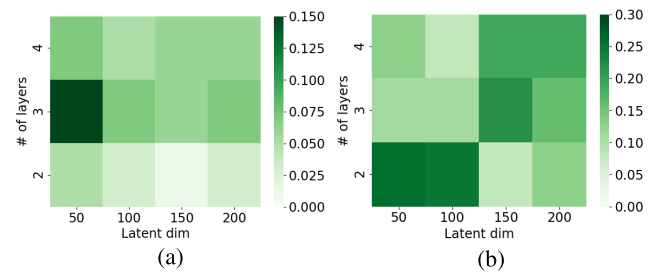
In addition, DCT is applied on the original  $80 \times 80$  image so that, 54 DCT coefficients were selected similarly and used for clustering. The results are compared with the proposed local patching strategy. Table 1 presents the average testing accuracies of the DCT analysis over the 4-subjects LOOCV for the original image and the selected local patch.

### B. ENTROPY ANALYSIS RESULTS

For the entropy analysis, the three patching strategies depicted in Fig. 7 are considered. The first two strategies resulted into 2-dimensional features, while the last patching strategy resulted into 10-dimensional feature vectors.



**FIGURE 12.** Illustration of the Dunn's index for the three patching strategies used for feature extraction based on entropy analysis.



**FIGURE 13.** (a) CAE and (b) CVAE robustness evaluation using Dunn's index.

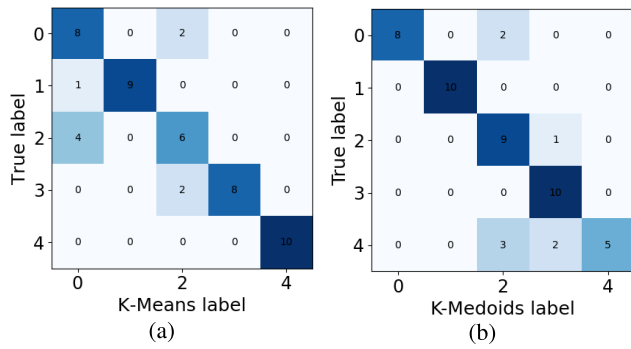
The results of the Dunn's indexes were computed using the K-Means clustering and presented in Fig. 12. As can be seen, the last patching strategy obtained the highest Dunn's index and therefore, it was selected for analysis. This result was expected, because the last strategy considers a higher number (10) of smaller patches. This represents local patterns variations better compared to the other two strategies. The other two strategies consider a fewer number (2) of larger local areas, which leads to poorer entropy computation.

### C. CAE AND CVAE ROBUSTNESS EVALUATION

The two deep NN architectures CAE and CVAE are evaluated addressing two criteria: 1) the number of hidden layers; and 2) the number of extracted features as latent dimension. In regards to the number of hidden layers, two, three, and four hidden layers are considered. The latent dimension is incorporated with the number of data samples in this study. The considered latent features are 50, 100, 150, and 200. Since the study is unsupervised, the true labels are seen as unknown. Hence, Dunn's index is selected for measuring the wellness of clusters separation. Fig. 13 reveals the Dunn's index for each experiment for CAE and CVAE.

Based on the results presented in Fig. 13a, three hidden layers with 50 extracted features are selected for CAE. In regards to the CVAE architecture, two hidden layers





**FIGURE 14.** Confusion matrices for comparison between local DCT-Based+K-Means and local DCT-Based+K-Medoids for different subjects.

with 50 extracted features are selected based on the results presented in Fig. 13b.

#### D. COMPARISON OF THE AVERAGE TRAINING AND TESTING ACCURACIES FOR K-MEANS AND K-MEDOIDS USING ALL FEATURE EXTRACTION TECHNIQUES

The average training and testing accuracies with standard deviations over 4-subjects LOOCV for K-Means and K-Medoids using all feature extraction methods are illustrated in Table 2 and Table 3.

As it can be observed, the two superior architectures are the local DCT-based frequency features extracted from the local patches and CVAE encoded features. After that, the local entropy analysis achieved the best results. K-Medoids has better performance than K-Means using the local DCT-based coefficient features. On the other hand, CVAE encoded features are better incorporated with K-Means. The results of the K-Means and K-Medoids are very similar in the case of local entropy-based features. In addition, CAE, PCA and 2DPCA have worse performance, while 2DPCA shows a minor improvement to PCA for K-Medoids.

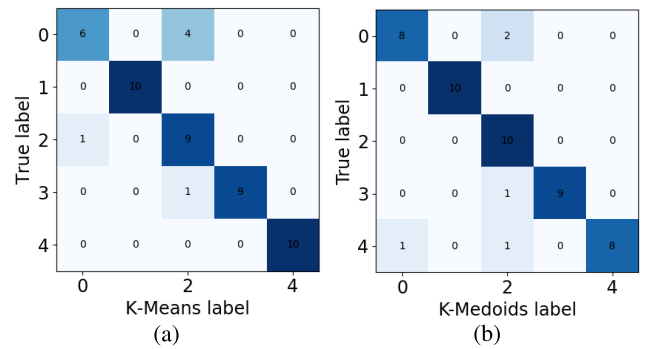
In order to evaluate the two superior architectures' performance for different activity groups, confusion matrices for K-Means and K-Medoids are visualised. The following matrices in Fig. 14 consider the proposed local DCT-based method over 4-subjects LOOCV for K-Means (Fig. 14a) and K-Medoids (Fig. 14b).

The confusion matrices for the CVAE encoded features with K-Means clustering and K-Medoids clustering are visualised in Fig. 15a and Fig. 15b respectively.

Both feature extraction strategies results show confusion of some activities. Further analysis of the results will be presented in the discussion section.

#### E. VISUALISATION RESULTS

Visualization of the raw data and CVAE encoded features are performed using t-SNE, MDS and LLE methods. Here, the actual data labels are used to map the samples. Initially, the original data  $D_{n \times m}$ , where  $m = 6400$ , is transformed and visualised in a 2-dimensional space as shown in Fig. 16. As illustrated, t-SNE performs reasonable



**FIGURE 15.** Confusion matrices for CVAE encoded features using K-Means (a) and K-Medoids (b) for different subjects.

separability between the classes, while there are some overlapping clusters in the case of MDS and LLE.

In order to perform comparison, the second scenario of data visualisation is concerned by transforming the encoded data  $D_{n \times 50}$  using t-SNE, MDS and LLE to a 2-dimensional space as seen in Fig. 17. Considering the illustration, all three methods showcase improvements in terms of cluster separability. However, there is still overlapping between the clusters. Since dimension reduction from 50 encoded features into only two features is a significant reduction in the number of features, no accurate clustering is expected using these two dimensional features. This is tested by applying K-Means and K-Medoids on the two t-SNE features. The average testing accuracies are 42% and 45% for K-Means and K-Medoids respectively. Hence, these manifold learning methods are good for visualisation, but are not necessarily accurate for clustering.

#### IV. DISCUSSION

The local patching strategy incorporated with DCT improved the average training and testing accuracies by 10%-15% for the two scenarios as seen in Table 1. It has been solely validated using an unsupervised metric for the scope of this study. As such, the method can be applied for other supervised or unsupervised studies with Doppler radar data. The local patching strategies can even be improved when used in a supervised framework, because the average validation accuracies allow optimum estimation of the method's parameters.

The confusion matrices for the architecture local DCT-Based+K-Means and local DCT-Based+K-Medoids in Fig. 14 reveal that the activities walking (1) and jumping (3) are problematic as they are frequently confused. Similarly, the CVAE-based architecture in Fig. 15 shows confusion between walking (1) and jumping (3). As it can be observed from the confusion matrices, the walking (1), jumping (3) and standing (5) classes are seen as problematic. More data can be collected in order to improve the results with a higher number of subjects. Considering the clustering methods, local DCT features are better incorporated with K-Medoids. On the other hand, the CVAE encoded features are more correctly clustered with K-Means. In addition, the results of

**TABLE 2.** Average training accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV.

	DCT-Based Method	Entropy-Based Method	CVAE	CAE	PCA	2DPCA
K-Means	80%±4.41	69.5%±3.23	83.75%±3.11	52%±4.74	64.25%±7.32	67%±10.29
K-Medoids	77.5%±8.87	69.5%±4.05	79.75%±3.11	54.5%±4.55	47%±4.41	64.5%±5.93

**TABLE 3.** Average testing accuracies with standard deviations of K-Means and K-Medoids using the five feature extraction methods over 4-subjects LOOCV.

	DCT-Based Method	Entropy-Based Method	CVAE	CAE	PCA	2DPCA
K-Means	75%±5.74	72%±7.11	84%±5.09	66%±8.6	58.5%±1.65	57.5%±2.59
K-Medoids	77%±4.58	72%±8.48	82%±4.89	65%±5.74	57%±2.23	64%±8.12

**TABLE 4.** Total computational time over 600 samples of the four folds for the two proposed feature extraction methods as well as CVAE.

	DCT-Based Method	Entropy-Based Method	CVAE
Time	0.58 s	1.1074 s	698.39 s

both clustering strategies were similar in the case of local entropy-based features. That is due to the lower resolution of the entropy features compared to DCT and CVAE techniques. In terms of efficiency, K-Means is shown to execute faster than K-Medoids [70].

Considering the employed feature extraction architectures, it can be concluded that the encoded data with CVAE is superior in comparison with the proposed local DCT-based and local entropy-based feature extraction methods in terms of accuracy. On the other hand, the local DCT-based and local entropy-based analyses are less complex and more easily implemented. The computational time for local DCT-based method and local entropy-based method is noticeably smaller in comparison with CVAE as seen in Table 4.

The reason is that, the local DCT-based and local entropy-based features are simply derived from the local patches. On the other hand, CVAE encoded features are mainly the results of convolution of the cubic Doppler images of size  $2 \times 100 \times 32$  with several filter types. Furthermore, the convolutional filters weights of both encoding and decoding structure are learnt based on an optimisation process using the objective function, which is computationally more complex rather than the other proposed strategies. Then, learning the local DCT-based and local entropy-based features from new coming datasets will be faster than the CVAE features. Given the advantages and disadvantages of each strategy, the overall recognition performances of them are reasonable and can be applied to other unsupervised project scenarios. PCA and CAE are not seen as successful due to their limitations. In regards to CAE, the method does not

include a regularization term, which is prone to producing inappropriate representations of the data samples in terms of classification. PCA preserves the global structure of the data, but fails to retain local dependencies in the lower dimensional space.

Considering the described individual feature extraction methods, one possible idea is to fuse the features from different strategies. However, there are reasons not to consider that for the current dataset. Given the limited number of samples in this study compared to the high number of features from most techniques, this will increase the dimensionality of the feature space. Therefore, it does not improve the accuracy. That is tested for the fusion of the local DCT-based features and local entropy-based features and no improvement was observed. In addition, since the clustering strategies are based on computation of the features distance and data fusion might require normalization of the heterogeneous feature types, that can also influence the clustering results.

The average training accuracies and average testing accuracies report a slightly bigger standard deviation in some cases. This is explained by the fact that 4-subject LOOCV is applied. As such, in some of the folds, the retained subject data for testing appears very different from the subjects data for training. However, the achieved accuracies are still reasonable for an unsupervised framework. That is a positive sign for the potential use of such strategies for e-Healthcare purposes.

The three manifold learning methods considered in this study can be extremely useful for data visualisation problems. As seen in Fig. 16, the t-SNE visualisation is better in comparison with MDS and LLE. Since the unsupervised CVAE-based architecture is known to provide a reasonable separation between the clusters, it is compared against the raw data. The results show better separation for all three visualisation techniques shown in Fig. 17. It can be concluded that the proposed deep CVAE improves the separability between the classes, which can boost the visualisation results for different manifold learning methods. For high-dimensional data

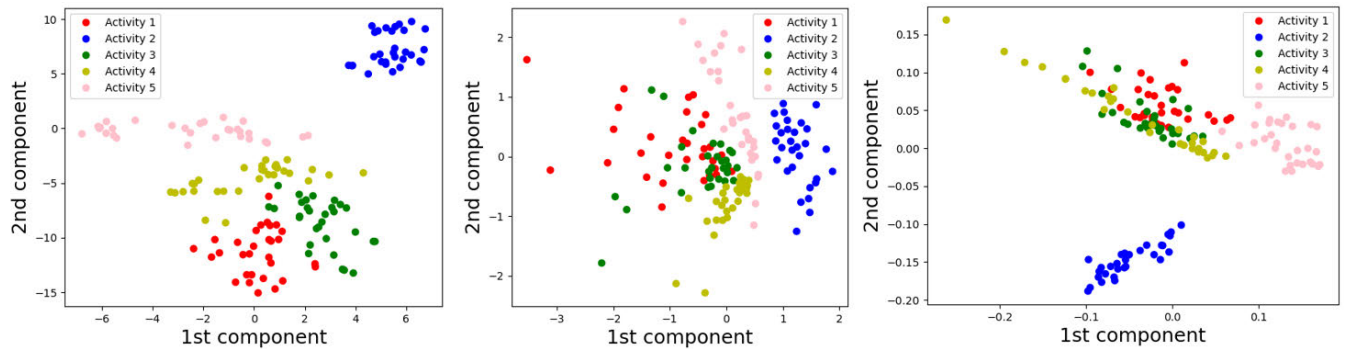


FIGURE 16. Visualisation of the transformed raw data  $D_{n \times 6400}$  using t-SNE, MDS, and LLE.

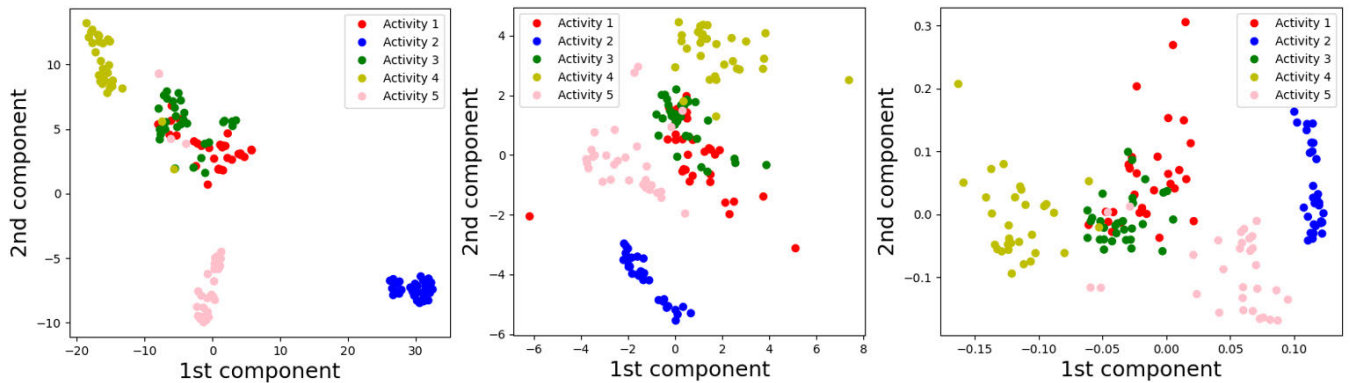


FIGURE 17. Visualisation of the transformed encoded data  $D_{n \times 50}$  using t-SNE, MDS, and LLE.

visualisation purposes, the CVAE encoded features can be used prior manifold learning. Despite the successful separation of classes, it can be observed that walking (1) and jumping (3) have overlapping samples. This is also observed in the confusion matrices in Section III, where these two activities are commonly mis-classified. Additionally, the samples from Activity 2 (running) are the most accurately separated from the remaining samples from the other activities. This finding is also evident when the manifold learning methods are applied on the raw data.

The achieved results are comparable with the previous supervised research framework on the same dataset in [20]. In that work, SVD, PCA and physical features are used for feature extraction. The average testing results of all three feature extraction methods combined with SVM for classification was reported to be more than 80% for different sizes of the training set. Similar result is observed with the proposed unsupervised local DCT-based method, and CVAE, where the average testing accuracies are more than 80% for the 4-subject LOOCV.

## V. CONCLUSION

This work studies the employment of Doppler radar for daily activity recognition using an unsupervised framework. The results of this study push the applications of Doppler radar data in healthcare one step forward to practice by enabling the recognition capability without label or with

poor labelling. In particular, the analysis architecture includes unsupervised feature extraction followed by clustering strategies. Four different categories of unsupervised feature extraction, namely, digital image frequency analysis based on DCT, entropy analysis, convolutional filtering based on deep autoencoders architectures CAE and CVAE, and the state of the art PCA and 2DPCA techniques, were employed. More specifically, two unsupervised methods for extraction of local DCT-based and local entropy-based features were proposed. The proposed two local patches-based methods for feature extraction exhibited an improvement of 5%-10% average testing accuracies compared to conventional CAE, PCA, and 2DPCA. On the other hand, the CVAE encoded features were superior with average testing accuracies of 84% and 82% for K-Means and K-Medoids respectively. Considering the expensive computational time for CVAE, the two proposed local DCT-based and local entropy-based methods provide a reasonable trade-off between time and accuracy. Regarding the unsupervised scenario, that is a positive sign for the potential use of these proposed techniques.

Three manifold learning methods for high-dimensional data visualisation are considered in this study - t-SNE, MDS and LLE. Visualization of the features using these three methods are compared. The results revealed that the clusters have a better separation with all three methods when the visualisation was performed on the CVAE encoded data. Finally, this project can serve as a reasonable application

with two proposed unsupervised feature extraction methods and a visualisation framework for project scenarios with poor labeling for both activity clustering and data visualisation.

## ACKNOWLEDGMENT

Yordanka Karayaneva would like to thank the sponsorship of the Data Driven Research Innovation (DDRI) at Coventry University, U.K.

## REFERENCES

- [1] *Living Longer—Office for National Statistics*, Office Nat. Statist., London, U.K., Aug. 2020.
- [2] S. U. Park, J. H. Park, M. A. Al-Masni, M. A. Al-Antari, M. Z. Uddin, and T.-S. Kim, "A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services," *Procedia Comput. Sci.*, vol. 100, pp. 78–84, Jan. 2016.
- [3] K.-T. Song and W.-J. Chen, "Human activity recognition using a mobile camera," in *Proc. 8th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Incheon, South Korea, Nov. 2011, pp. 3–8, doi: [10.1109/URAI.2011.6145923](https://doi.org/10.1109/URAI.2011.6145923).
- [4] A. Jalal, S. Kamal, and D. Kim, "A depth video-based human detection and activity recognition using multi-features and embedded hidden Markov models for health care monitoring systems," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, no. 4, p. 54, 2017.
- [5] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 9, no. 21, pp. 1–17, 2012.
- [6] M. Al-Khafajiy, T. Baker, C. Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich, "Remote health monitoring of elderly through wearable sensors," *Multimedia Tools Appl.*, vol. 78, pp. 1–26, Jan. 2019.
- [7] J. Liu, J. Sohn, and S. Kim, "Classification of daily activities for the elderly using wearable sensors," *J. Healthcare Eng.*, vol. 2017, Nov. 2017, Art. no. 8934816.
- [8] Z. Wang, Z. Yang, and T. Dong, "A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time," *Sensors*, vol. 17, no. 2, p. 341, 2017, doi: [10.3390/s17020341](https://doi.org/10.3390/s17020341).
- [9] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, J. Wang, and J. Kim, "Wearable sensors: Modalities, challenges, and prospects," *Lab Chip*, vol. 18, no. 2, pp. 217–248, 2018, doi: [10.1039/c7lc00914c](https://doi.org/10.1039/c7lc00914c).
- [10] A. Kamišalić, I. Fister, M. Turkanović, and S. Karakatić, "Sensors and functionalities of non-invasive wrist-wearable devices: A review," *Sensors*, vol. 18, no. 6, p. 1714, May 2018, doi: [10.3390/s18061714](https://doi.org/10.3390/s18061714).
- [11] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, p. 21, 2012.
- [12] J. Huberty, D. K. Ehlers, J. Kurka, B. Ainsworth, and M. Buman, "Feasibility of three wearable sensors for 24 hour monitoring in middle-aged women," *BMC Women's Health*, vol. 15, no. 1, p. 15, Dec. 2015.
- [13] Y. Karayaneva, S. Baker, B. Tan, and Y. Jing, "Use of low-resolution infrared pixel array for passive human motion movement and recognition," in *Proc. HCI*, Belfast, U.K., 2018, pp. 1–5.
- [14] S. Mashiyama, J. Hong, and T. Ohtsuki, "Activity recognition using low resolution infrared array sensor," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 495–500.
- [15] S. Mashiyama, J. Hong, and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," in *Proc. IEEE 25th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Washington, DC, USA, Sep. 2014, pp. 2109–2113.
- [16] Y. Karayaneva, S. Sharifzadeh, Y. Jing, and B. Tan, "Infrared human activity recognition dataset—Coventry-2018," IEEE Dataport, 2020, doi: [10.21227/baja-1j59](https://doi.org/10.21227/baja-1j59).
- [17] H. Chen and W. Ye, "Classification of human activity based on radar signal using 1-D convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1178–1182, Jul. 2020, doi: [10.1109/LGRS.2019.2942097](https://doi.org/10.1109/LGRS.2019.2942097).
- [18] Y. He, Y. Yang, Y. Lang, D. Huang, X. Jing, and C. Hou, "Deep learning based human activity classification in radar micro-Doppler image," in *Proc. 15th Eur. Radar Conf. (EuRAD)*, Madrid, Spain, Sep. 2018, pp. 230–233, doi: [10.23919/EuRAD.2018.8546615](https://doi.org/10.23919/EuRAD.2018.8546615).
- [19] Q. Chen, B. Tan, K. Woodbridge, and K. Chetty, "Doppler based detection of multiple targets in passive Wi-Fi radar using underdetermined blind source separation," in *Proc. Int. Conf. Radar (RADAR)*, Brisbane, QLD, Australia, Aug. 2018, pp. 1–6.
- [20] W. Li, B. Tan, and R. Piechocki, "Passive radar for opportunistic monitoring in E-health applications," *IEEE J. Transl. Eng. Health Med.*, vol. 6, Jan. 2018, Art. no. 2800210, doi: [10.1109/JTEHM.2018.2791609](https://doi.org/10.1109/JTEHM.2018.2791609).
- [21] F. Wang, M. Skubic, M. Rantz, and P. E. Cuddihy, "Quantitative gait measurement with pulse-Doppler radar for passive in-home gait assessment," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2434–2443, Sep. 2014, doi: [10.1109/TBME.2014.2319333](https://doi.org/10.1109/TBME.2014.2319333).
- [22] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. ACM Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 95–108.
- [23] W. Li, B. Tan, Y. Xu, and R. J. Piechocki, "Log-likelihood clustering-enabled passive RF sensing for residential activity recognition," *IEEE Sensors J.*, vol. 18, no. 13, pp. 5413–5421, Jul. 2018, doi: [10.1109/JSEN.2018.2834739](https://doi.org/10.1109/JSEN.2018.2834739).
- [24] G. Diraco, A. Leone, and P. Siciliano, "A radar-based smart sensor for unobtrusive elderly monitoring in ambient assisted living applications," *Biosensors*, vol. 7, no. 4, p. 55, Nov. 2017, doi: [10.3390/bios7040055](https://doi.org/10.3390/bios7040055).
- [25] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, May 2009.
- [26] F. Fioranelli, M. Ritchie, S. Z. Gurbuz, and H. Griffiths, "Feature diversity for optimized human micro-Doppler classification using multistatic radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 640–654, Apr. 2017.
- [27] P. P. Markopoulos and F. Ahmad, "Indoor human motion classification by L1-norm subspaces of micro-Doppler signatures," in *Proc. IEEE Radar Conf. (RadarConf)*, Seattle, WA, USA, May 2017, pp. 1807–1810.
- [28] A.-K. Seifert, L. Schäfer, M. G. Amin, and A. M. Zoubir, "Subspace classification of human gait using radar micro-Doppler signatures," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Rome, Italy, Sep. 2018, pp. 311–315.
- [29] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 6, pp. 3617–3628, Dec. 2019.
- [30] W. Li, Y. Xu, B. Tan, and R. J. Piechocki, "Passive wireless sensing for unsupervised human activity recognition in healthcare," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Valencia, Spain, Jun. 2017, pp. 1528–1533, doi: [10.1109/IWCMC.2017.7986511](https://doi.org/10.1109/IWCMC.2017.7986511).
- [31] M. S. Seyfioglu, A. M. Ozbayoglu, and S. Z. Gurbuz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018.
- [32] Y. H. Liu, "Feature extraction and image recognition with convolutional neural networks," *J. Phys., Conf. Ser.*, vol. 1087, Sep. 2018, Art. no. 062032.
- [33] F. Zhu, L. Shao, J. Xie, and Y. Fang, "From handcrafted to learned representations for human action recognition: A survey," *Image Vis. Comput.*, vol. 55, pp. 42–52, Nov. 2016.
- [34] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, "Convolutional learning of spatio-temporal features," in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Germany: Springer, 2010, pp. 140–153.
- [35] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
- [36] Y. Wu, P. Tamayo, and K. Zhang, "Visualizing and interpreting single-cell gene expression datasets with similarity weighted nonnegative embedding," *Cell Syst.*, vol. 7, pp. 656–666, Dec. 2018.
- [37] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: [10.1007/BF02289263](https://doi.org/10.1007/BF02289263).
- [38] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J-Multidisciplinary Sci. J.*, vol. 2, no. 2, pp. 226–235, Jun. 2019.
- [39] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
- [40] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
- [41] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, Jan. 1974.



- [42] C. Karabacak, S. Z. Gurbuz, A. C. Gurbuz, M. B. Guldogan, G. Hendeby, and F. Gustafsson, "Knowledge exploitation for human micro-Doppler classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2125–2129, Oct. 2015, doi: [10.1109/LGRS.2015.2452311](https://doi.org/10.1109/LGRS.2015.2452311).
- [43] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016, doi: [10.1109/LGRS.2015.2491329](https://doi.org/10.1109/LGRS.2015.2491329).
- [44] H. Du, T. Jin, Y. Song, Y. Dai, and M. Li, "A three-dimensional deep learning framework for human behavior analysis using range-Doppler time points," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 611–615, Apr. 2020, doi: [10.1109/LGRS.2019.2930636](https://doi.org/10.1109/LGRS.2019.2930636).
- [45] M. R. I. Minto, B. Tan, S. Sharifzadeh, T. Riihonen, and M. Valkama, "Shallow neural networks for mmWave radar based recognition of vulnerable road users," in *Proc. 12th Int. Symp. Commun. Syst., Netw. Digit. Signal Process. (CSNDSP)*, Porto, Portugal, Jul. 2020, pp. 1–6, doi: [10.1109/CSNDSP49049.2020.9249537](https://doi.org/10.1109/CSNDSP49049.2020.9249537).
- [46] Math Works. (2019). *Discrete Cosine Transform-MATLAB & Simulink*. [Online]. Available: <https://www.mathworks.com/help/images>
- [47] P. Molchanov, J. Astola, K. Egiazarian, and A. Totsky, "Ground moving target classification by using DCT coefficients extracted from micro-Doppler radar signatures and artificial neuron network," in *Proc. Microw., RADAR REMOTE Sens. Symp.*, Kiev, Ukraine, Aug. 2011, pp. 173–176.
- [48] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL, USA: Univ. Illinois Press, 1949.
- [49] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. ICML Workshop Unsupervised Transf. Learn.*, vol. 27, 2012, pp. 37–50.
- [50] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
- [51] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini, "Anomaly detection using autoencoders in high performance computing systems," 2018, *arXiv:1811.05269*. [Online]. Available: <https://arxiv.org/abs/1811.05269>
- [52] W. Xu, K. Shawn, and G. Wang, "Adversarially approximated autoencoder for image generation and manipulation," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2387–2396, Sep. 2019.
- [53] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3581–3589.
- [54] C. Dong, T. Xue, and C. Wang, "The feature representation ability of variational autoencoder," in *Proc. IEEE 3rd Int. Conf. Data Sci. Cyberspace (DSC)*, Guangzhou, China, Jun. 2018, pp. 680–684, doi: [10.1109/DSC.2018.00108](https://doi.org/10.1109/DSC.2018.00108).
- [55] H. Nishizaki, "Data augmentation and feature extraction using variational autoencoder for acoustic modeling," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPAASC)*, Dec. 2017, pp. 1222–1227.
- [56] P. A. Andersen, M. Goodwin, and O. C. Granmo, "The dreaming variational autoencoder for reinforcement learning environments," in *Artificial Intelligence XXXV*, vol. 11311, M. Bramer and M. Petridis, Eds. Cham, Switzerland: Springer, Dec. 2018, pp. 143–155. [Online]. Available: [http://link.springer.com/10.1007/978-3-030-04191-5\\_11](http://link.springer.com/10.1007/978-3-030-04191-5_11), doi: [10.1007/978-3-030-04191-5\\_11](https://doi.org/10.1007/978-3-030-04191-5_11).
- [57] B. Erol, S. Z. Gurbuz, and M. G. Amin, "Motion classification using kinematically sifted acgan-synthesized radar micro-doppler signatures," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 4, pp. 3197–3213, Aug. 2020.
- [58] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019.
- [59] C. Doersch, "Tutorial on variational autoencoders," 2016, *arXiv:1606.05908*. [Online]. Available: <https://arxiv.org/abs/1606.05908>
- [60] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.
- [61] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
- [62] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2013, pp. 423–430.
- [63] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009, pp. 515–518.
- [64] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [65] B. M. Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Sci. Int.*, vol. 311, Jun. 2020, Art. no. 110194.
- [66] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1–27, Mar. 1964.
- [67] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, Jun. 1964.
- [68] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009, pp. 570–572.
- [69] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [70] T. Velmurugan, "Efficiency of K-means and K-medoids algorithms for clustering arbitrary data points," *Int. J. Comput. Technol. Appl.*, vol. 3, no. 5, pp. 1758–1764, 2012.



**YORDANKA KARAYANEVA** received the B.Sc. degree in computer science from Coventry University, U.K., in 2017, where she is currently pursuing the Ph.D. degree in computer science.

She was a Research Intern in reinforcement learning (from June 2016 to July 2016) and in distributed systems (from June 2017 to August 2017) with Coventry University. She was a Ph.D. Teaching Assistant with Coventry University, from September 2018 to April 2019. She has published conference papers in IEEE and ACM venues. Her research interests include machine learning, signal processing, object recognition, and computer vision.



**SARA SHARIFZADEH** received the B.Sc. and M.Sc. degrees in electronics engineering from the University of Mazandaran, Iran, in 2003 and 2007, respectively, the second M.Sc. degree in multimedia technologies from the Universitat Autònoma de Barcelona, Spain, in 2010, and the Ph.D. degree in computer science from the Technical University of Denmark, in 2015.

From 2015 to 2017, she was a Postdoctoral Research Associate in data science with Loughborough University, U.K. Since 2018, she has been a Lecturer with Coventry University. She has published in several leading conferences and journals. Her research interests include machine learning, artificial intelligence, their application on analysis of digital signals, images, and 3D point clouds, signal and image analysis for gesture recognition, monitoring, and remote sensing applications.



**WENDA LI** received the M.Eng. and Ph.D. degrees from the University of Bristol, in 2013 and 2017, respectively. Before joining University College London, he worked with the University of Birmingham, as a Research Fellow. He is currently a Research Fellow with the Department of Security and Crime Science, University College London. His research interest includes signal processing for passive radar and high-speed digital system design for wireless sensing applications in healthcare, security, and positioning. His research in passive WiFi radar has led to a number of IEEE conference and journal publications.





**YANGUO JING** (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science from Dalian Maritime University, China, in 1997 and 2000, respectively, and the Ph.D. degree in computer science from Heriot-Watt University, U.K., in 2004.

From 2002 to 2016, he was an Associate Professor with London Metropolitan University, U.K. Since 2016, he has been a Professor with Coventry University, U.K. His research interests include

user modeling, artificial intelligent, machine learning, and machine learning methods in intelligent applications. He is currently a member of IET and a Fellow of BCS and CITP. He was a recipient of the Best Paper Award in the 11th International Conference on Developments in eSystems Engineering (DeSE), in 2018.



**BO TAN** (Member, IEEE) received the B.Sc. and M.Sc. degrees in communications engineering from the Beijing University of Posts and Telecommunications, in 2004 and 2008, respectively, and the Ph.D. degree from the Institute for Digital Communications, The University of Edinburgh, U.K., in 2013.

From 2012 to 2016, he was a Postdoctoral Research Associate with University College London and the University of Bristol. From 2017 to

2018, he was a Lecturer with Coventry University. Since 2019, he has been a tenure track Assistant Professor with Tampere University, Finland. His research has led to 50 academic publications and U.S. patent. His research interests include radio signal processing in radar, wireless communications systems, machine learning methods of sensing data, machine learning enabled wireless sensing, and joint radar-communications design in intelligent machines. He is a member of ACM. He is also an active reviewer of multiple IEEE and IET journals in sensing and communications.

...